

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANTS: Aaron Kaplan et al.
- (ii) TITLE OF INVENTION: ENHANCING INORGANIC CARBON FIXATION BY PHOTOSYNTHETIC ORGANISMS
- (iii) NUMBER OF SEQUENCES: 9
- (iv) CORRESPONDENCE ADDRESS:
- (A) ADDRESSEE: Mark M. Friedman c/o Anthony Castorina
- (B) STREET: 2001 Jefferson Davis Highway, Suite 207
- (C) CITY: Arlington
- (D) STATE: Virginia
- (E) COUNTRY: United States of America
- (F) ZIP: 22202
- (v) COMPUTER READABLE FORM:
- (A) MEDIUM TYPE: 1.44 megabyte, 3.5" microdisk
- (B) COMPUTER: Twinhead Slimnote-890TX
- (C) OPERATING SYSTEM: MS DOS version 6.2,
Windows version 3.11
- (D) SOFTWARE: Word for Windows version 2.0 converted to
an ASCII file
- (vi) CURRENT APPLICATION DATA:
- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
- (A) APPLICATION NUMBER:
- (B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
- (A) NAME: Friedmam, Mark M.
- (B) REGISTRATION NUMBER: 33,883
- (C) REFERENCE/DOCKET NUMBER: 325/45
- (ix) TELECOMMUNICATION INFORMATION:
- (A) TELEPHONE: 972-3-5625553
- (B) TELEFAX: 972-3-5625554
- (C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4957
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

AAGCTTGGAT TGAAGCGATC GGGGTCAATC CCAGCGATGA TCCTCAGTTC   50
CTCCTGATGG TCGATCCCTT TAGCGCCAAG ATTGAGGATC TGCTGCAAGG  100
GCTGGATTTC GCCTATCCCG AGGCCGTGAA AGTGGGCGGA TTGGCCAGTG  150
GTTTGGGGGC AGAGTCAGCG ATCGCCAGCT TGTTTTTCA AGACCGACAG  200
GTCGATGGCG TGATTGGGCT AGCCCTCAGT GGCAATGTCC AGCTGCAGGC  250
GATCGTGGCT CAGGGCTGTC GTCCAGTTGG CCCGCTTGG CATGTGGCAG  300
CGGCGGAGCG CAACATCTCTG CGGCAACTTC AGACCGAAGA CGAGGAACCG  350
ATCGCCGCGC TGCAAGCCCT ACAGTCAGTC CTGCGTGATC TCTCCCTGA  400
ATTACAGCGA TCGCTCTGTG TGGGCTTGGC CTGCAATTCT TTCAAACGG  450
TATTACAACC GGGCGACTTC CTGATCCGTA ACCTGCTGGG GTTTGATCCC  500
CGCACTGGTG CTGTAGCAAT CGGCGATCGC ATTCGAGTTG GGCAGCGGCT  550
GCAGCTGCAC GTACGGGATG CCCAGACAGC GGCGGATGAC CTCGAGCGGC  600
AACTGGGGCA ATGGTGCCGG CAGCATGCGA CAAAACCAGC AGCTTCCCTC  650
TTGTTTTCTT GCTTGGGGCG CGGCAAGCCC TTCTATCAGC AGGCCAACTT  700

```

CGAGTCGCAA CTGATTCAGC ATTACCTCTC AGAGCTGCCC CTAGCTGGCT 750
 TTTTCTGTAA TGGCGAAATC GGCCCGATCG CTGGCAGCAC CTACCTGCAT 800
 GGCTACACAT CGGTGCTGGC TTTGCTGTCT GCCAAACTC ACTAGCGCCA 850
 GCGAGACCTG ATTGTCGATC TGCTGAGCGC GACTGTAGCG CTGGAAATAG 900
 GCGCGACCT GAGCAGGCGC ATCGGCCAAG CTGACCGTAG TATCACCGTC 950
 AGCCACCCCC GCCAGAAAT TCCGCAACAT CGGCAGGAGA GCGATCGCCT 1000
 CCGCCTCCGA TAAATTCAAC GGCTCATGGG TCAACAGGCG GATCAAGTAC 1050
 TCTGACTGCG ATCGCCATCC ATTCCCGCCG AAAACGTTTG TAAATCAGTC 1100
 TTGATCCGGT AGCGATCGCA CCCGACGGGA CTCTAGTTCT AGTTGCCAAC 1150
 CTTCAGCGGC AGGTTGTACG GTTCCGAGTC GGTAGGGATG GGGATAGCTG 1200
 ACCAAGGAAC CGGTCGTGAC TTCCAGAGA GCACCTTGCT GACTGGTGGC 1250
 TTGGATGTGG AGGTGGCCTG TGAAGATCAC CGAGACGCTG CCCGCTTCGA 1300
 GGATTGATCG CAATTCCTCG GCATTTTCTA AGATGTAGCG CTGACCAAGC 1350
 GGATGCTGCT GTTGATCGGG CAGATGCTCC AACACATTGT GGTGAATCAT 1400
 CACCCAGCGT TGGCTAGCGG TGGAAAGTGGC GAGTTCTTGT TGCAGCCAGT 1450
 TGAGTTGCGC GCAATCGACT CGCCCCGAT GCAGTTGATG GCCCGCTTCA 1500
 TCAAAAGCGA TCGAATTGAG CGCAAACAGA TCGAGATCCG GTGCGATCGT 1550
 GCAGCGATAG TAGGGGCGAT CGCTCGTGAA GCCAAAGTCT TGATAGAGCT 1600
 CGACAAACTC GGCCACACCG GTGCGATCGC GATCGCTCGC TGCGGCGGGC 1650
 ATATCGTGGT TGCCCGGCAC CACATAGACC GGATAGGGCA ACTGGCGCAA 1700
 TTGTTGCAGC AGCCACTGAT GGTTTTCCCG CTCCCGTGC TGGGTTAAAT 1750
 CCCCCGGCAG CAACAGGAAG TCCAAATCCA GCGCTGCCAG TTCTGTCAGG 1800
 ATTTGCTCAA AAGCCGGAAT GCTGCACTCA ATCAAATGGA AGCGATGGGG 1850
 ATGGTGCCAA ATTGTCTGCG GCAGTCCAAT GTGGAGATCG CTCAGCAGCG 1900
 CAAATCGAAA CGCTCGGTTT ATTGCCATCC CCTCAGCTAT CGAGCCCGAT 1950
 TCTAGGCGAA GCTAGGTCGA GTCCGTGTGC TTCAGTTGCA AGCATTTCATG 2000
 GCCAGAGTTC GCGTTCGGCA GCACGTCAAT CCGCTCTCTC AGAAATTCGA 2050
 AGTGGTCACG ACTTGGCCCG ATTGGCAACA GGTCTATGCG GACTGCGATC 2100
 GCGCGCTGCA TTTGGATATT GGCTGTGCTC GCGGGCGCTT TCTGCTGGCA 2150
 ATGGCGACAC GACAACTTGA GTGGAATTAT CTGGGGCTGG AAATTCGTGA 2200
 GCGCTGGTA GATGAGGCGA ACGCGATCGC CCGCGAACGT GAACTGACCA 2250
 ATCTCTACTA CCACTTCAGC AACGCCAATT TGGACTTGGA ACCGCTGCTG 2300
 CGATCGCTGC CGACAGGGAT TTTGCAGCGG GTCAGCATTC AGTTCCCGGA 2350
 TCCTTGGTTC AAGAAACGCC ATCAAAGCG ACGCGTCGTC CAGCCGGAAC 2400
 TGGTGCAAGC CCTCGGACT GCGTTACCTG CTGGTGCAAG GGTCTTTCTG 2450
 CAATCCGATG TGCTGGAAGT GCAGGCAGAG ATGTGCGAAC ACTTTGCGGC 2500
 GGAACCCCGC TTTAGCGCA CTGCTTGGA CTGGCTGCCG GAAATCCGC 2550
 TGCCCCTCCC GACCGAGCGC GAAATTGCCG TTCAAACAA ACAGTTGCCA 2600
 GTCTACCGTG CTCTCTTCAT TCGGCAGCCA GCGGACTAAG CTCTTAAGGC 2650
 AAGCGTTGAC GCGATCGCGA TGAATGTCTG GCAAACCTG ACTTTTGCCC 2700
 ATTACCAACC CCAACAGTGG GGCCACAGCA GTTCTTGCA TCGGCTGTTT 2750
 GGCAGCCTGC GAGCTTGGC GGCCTCCAGC CAGCTGTTGG TTTGGTCTGA 2800
 GGCACTGGGT GGCTTCTTGC TTGCTGTCGT CTACGGTTG GTCCTGTTG 2850
 TGCCCAAGTC CGCCCTAGGA TTGGGGCTAG CCGCATCGC GGCTATTGG 2900
 GCCCTGCTCT CGCTGACAGA TATCGATCTG CGGCAAGCAA CCCCATTCA 2950
 CTGGCTGGTG CTGCTCTACT GGGGCGTCGA TGCCCTAGCA ACGGGACTCT 3000
 CACCCGTACG CGCTGCAGCT TTAGTTGGGC TAGCCAAACT GACGCTCTAC 3050
 CTGTTGGTTT TTGCCCTAGC GGCTCGGGT CTCCGCAATC CCCGTCTGCG 3100
 ATCGCTGCTG TTCTCGGTCG TCGTGATCAC ATCGCTTTT GTCACTGTCT 3150
 ACGGCCTCAA CCAATGGATC TACGGCGTTG AAGAGCTGGC GACTTGGGTG 3200
 GATCGCAACT CGGTTGCCGA CTTACCTCA CGGGTTTACA GCTATCTGGG 3250
 CAACCCCAAC CTGCTGGCTG CTTATCTGGT GCCGACGACT GCCTTTTCTG 3300
 CAGCAGCGAT CGGGGTGTGG CGCGGCTGGC TCCCAAGCT GCTGGCGATC 3350
 GCTGCGACAG GTGCGAGCAG CTTATGTCTG ATCCTCACCT ACAGTCGCGG 3400
 TGGCTGGCTG GGTTTTGTCT CCATGATTTT TGTCTGGGCG TTATTAGGGC 3450
 TCTACTGGTT TCAACCCCGT CTACCCGCAC CCTGGCGACG CTGGCTATTG 3500
 CCAATCGTAT TGGGTGGACT AGTCGCGGTG CTCTTGGTGG CGGTGCTTGG 3550

ACTTGAGCCG TTGCGCGTGC GCGTGTGAG CATCTTTGTG GGGCGTGAAG 3600
 ACAGCAGCAA CAACTTCCGG ATCAATGTCT GGCTGGCGGT GCTGCAGATG 3650
 ATTCAAGATC GGCCTTGGCT GGGCATCGGC CCCGGCAATA CCGCCTTTAA 3700
 CCTGGTTTAT CCCCTCTATC AACAGGCGCG CTTTACGGCG TTGAGCGCCT 3750
 ACTCCGTCCC GCTGGAAGTC GCGGTTGAGG GCGGACTACT GGGCTTGACG 3800
 GCCTTCGCTT GGCTGCTGCT GGTACGCGG GTGACGGCGG TCGGCGAGGT 3850
 GAGCCGACTG CGGCGCGATC GCAATCCCCA AGCCTTTTGG TTGATGGCTA 3900
 GCTTGGCCGG TTTGGCAGGA ATGCTGGGTC ACGGTCTGTT TGATACCGTG 3950
 CTCTATCGAC CGGAAGCCAG TACGCTCTGG TGGCTCTGTA TTGGAGCGAT 4000
 CGCGAGTTTC TGGCAGCCCC AACCTTCCAA GCAACTCCCT CCAGAAGCCG 4050
 AGCATTGAGA CGAAAAATG TAGCGGGCTC CCCAACAAT TCCTGTGCAC 4100
 CCGACTGGAT CCACCACCTA AACTGGATCC CAAAGGTATC CGGTGGATCT 4150
 AGGGTCATAA CGAACTCCGA CCGCGATCGC GTCCGCGAAC TGAACCTCCA 4200
 TCGCACCGAA GCGGAGTTCG TTAGTCGTTG AAGAGCCAAT GCTAGAGGGG 4250
 GCTGCCGAAG CAGTTGGGCT GGAAGCAGGC TCGGAGAAGC CACCCGCATC 4300
 CAAGGCAAG TTGAGCCGAC CTTCCGCAA GACTACGATC GCCACGGCGG 4350
 CTCTGCCAGC TAAGTCAGCG CTGGGTTAGT TGTCATAGCA GTCCGAGAC 4400
 AAGTTAGGAC AACTTCATAG AGGGACTCGC TCAGAGTCAA CAGCCGCTGT 4450
 CCGTGGGGGT GCGCAATCAC CCCACACCC ACGCACTGGG GGACTCGACT 4500
 CCCCCAGGCC CCCCACAACA AGATTTCGGA TAAGGGGCAT CGGCTGAATC 4550
 GCGATCGCTG CGGTAATAAC TAGCCGGTGT TAGCCATGGG TTTGAGACTA 4600
 ATCGGCACGG GGCAAAACGT CCTGATTAT TTGCTCAATG TGATAGGTTA 4650
 CATCGTCAAA AACAGGCCC AAGAGGTAGG AAAAATCAG ACCGCCAAG 4700
 TCCGAGGGCT TTGCTGTTGG GAGCGACCTA GGGCAGACTA GACAGAGCAT 4750
 TGCTGTGAGC CAAAGCGCCT TCAATTGCTG GCGGCTGTGG GTTTTTCGGA 4800
 GGTTGCCAAA TGAAGACCT TTTCGTCAAT GTCCTCCGCT ATCCCCGCTA 4850
 CTTTCATACC TTCCAGCTGG GTATTTTTTA GTCGATCTAC CAGTGGGTGC 4900
 GGCCGATGGT TCGCAACCCA GTCGCGGCTT GGGCGCTGCT AGGCTTTGGA 4950
 GTTTCGA 4957

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1404
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

ATGACTGTCT GGCAAACTCT GACTTTTGCC CATTACCAAC CCCAACAGTG 50
 GGGCCACAGC AGTTTCTTGC ATCGGCTGTT TGGCAGCCTG CGAGCTTGGC 100
 GGGCCTCCAG CCAGCTGTTG GTTTGGTCTG AGGCACTGGG TGGCTTCTTG 150
 CTTGCTGTCG TCTACGGTTC GGCTCCGTTT GTGCCAGTT CCGCCCTAGG 200
 GTTGGGGCTA GCCGCGATCG CGGCCTATTG GGCCCTGCTC TCGCTGACAG 250
 ATATCGATCT GCGGCAAGCA ACCCCCATTC ACTGGCTGGT GCTGCTCTAC 300
 TGGGGCGTCG ATGCCCTAGC AACGGGACTC TCACCCGTAC GCGCTGCAGC 350
 TTTAGTTGGG CTAGCCAAAC TGACGCTCTA CCTGTTGGTT TTTGCCCTAG 400
 CGGCTCGGGT TCTCCGCAAT CCCCGTCTGC GATCGCTGCT GTTCTCGGTC 450
 GTCGTGATCA CATCGCTTTT TGTCAGTGTC TACGGCCTCA ACCAATGGAT 500
 CTACGGCGTT GAAGAGCTGG CGACTTGGGT GGATCGCAAC TCGGTTGCCG 550
 ACTTCACCTC ACGGGTTTAC AGCTATCTGG GCAACCCCAA CCTGCTGGCT 600
 GCTTATCTGG TGCCGACGAC TGCCTTTTCT GCAGCAGCGA TCGGGGTGTG 650
 GCGCGGCTGG CTCCCCAAGC TGCTGGCGAT CGCTGCGACA GGTGCGAGCA 700
 GCTTATGTCT GATCCTCACC TACAGTCGCG GTGGCTGGCT GGTTTTGTG 750
 GCCATGATTT TTGTCTGGGC GTTATTAGGG CTCTACTGGT TTCAACCCCG 800
 TCTACCCGCA CCCTGGCGAC GCTGGCTATT CCCAGTCGTA TTGGGTGGAC 850
 TAGTCGCGGT GCTCTTGGTG GCGGTGCTTG GACTTGAGCC GTTGCAGCTG 900
 CGCGTGTTGA GCATCTTTGT GGGGCGTGAA GACAGCAGCA ACAACTTCCG 950
 GATCAATGTC TGCTGGCGG TGCTGCAGAT GATTCAAGAT CGGCCTTGGC 1000

TGGGCATCGG CCCCGGCAAT ACCGCCTTTA ACCTGGTTTA TCCCCTCTAT 1050
 CAACAGGCGC GCTTTACGGC GTTGAGCGCC TACTCCGTCC CGCTGGAAGT 1100
 CGCGGTTGAG GCGGGAATAC TGGGCTTGAC GGCCTTCGCT TGGCTGCTGC 1150
 TGGTCACGGC GGTGACGGCG GTGCGGCAGG TGAGCCGACT GCGGCGCGAT 1200
 CGCAATCCCC AAGCCTTTTG GTTGATGGCT AGCTTGGCCG GTTTGGCAGG 1250
 AATGCTGGGT CACGGTCTGT TTGATACCGT GCTCTATCGA CCGGAAGCCA 1300
 GTACGCTCTG GTGGCTCTGT ATTGGAGCGA TCGCGAGTTT CTGGCAGCCC 1350
 CAACCTTCCA AGCAACTCCC TCCAGAAGCC GAGCATTCAG ACGAAAAAAT 1400
 GTAG 1404

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 467
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Thr Val Trp Gln Thr Leu Thr Phe Ala His Tyr Gln Pro Gln
 5 10 15
 Gln Trp Gly His Ser Ser Phe Leu His Arg Leu Phe Gly Ser Leu
 20 25 30
 Arg Ala Trp Arg Ala Ser Ser Gln Leu Leu Val Trp Ser Glu Ala
 35 40 45
 Leu Gly Gly Phe Leu Leu Ala Val Val Tyr Gly Ser Ala Pro Phe
 50 55 60
 Val Pro Ser Ser Ala Leu Gly Leu Gly Leu Ala Ala Ile Ala Ala
 65 70 75
 Tyr Trp Ala Leu Leu Ser Leu Thr Asp Ile Asp Leu Arg Gln Ala
 80 85 90
 Thr Pro Ile His Trp Leu Val Leu Leu Tyr Trp Gly Val Asp Ala
 95 100 105
 Leu Ala Thr Gly Leu Ser Pro Val Arg Ala Ala Ala Leu Val Gly
 110 115 120
 Leu Ala Lys Leu Thr Leu Tyr Leu Leu Val Phe Ala Leu Ala Ala
 125 130 135
 Arg Val Leu Arg Asn Pro Arg Leu Arg Ser Leu Leu Phe Ser Val
 140 145 150
 Val Val Ile Thr Ser Leu Phe Val Ser Val Tyr Gly Leu Asn Gln
 155 160 165
 Trp Ile Tyr Gly Val Glu Glu Leu Ala Thr Trp Val Asp Arg Asn
 170 175 180
 Ser Val Ala Asp Phe Thr Ser Arg Val Tyr Ser Tyr Leu Gly Asn
 185 190 195
 Pro Asn Leu Leu Ala Ala Tyr Leu Val Pro Thr Thr Ala Phe Ser
 200 205 210
 Ala Ala Ala Ile Gly Val Trp Arg Gly Trp Leu Pro Lys Leu Leu
 215 220 225
 Ala Ile Ala Ala Thr Gly Ala Ser Ser Leu Cys Leu Ile Leu Thr
 230 235 240
 Tyr Ser Arg Gly Gly Trp Leu Gly Phe Val Ala Met Ile Phe Val
 245 250 255
 Trp Ala Leu Leu Gly Leu Tyr Trp Phe Gln Pro Arg Leu Pro Ala
 260 265 270
 Pro Trp Arg Arg Trp Leu Phe Pro Val Val Leu Gly Gly Leu Val
 275 280 285
 Ala Val Leu Leu Val Ala Val Leu Gly Leu Glu Pro Leu Arg Val
 290 295 300

Arg Val Leu Ser Ile Phe Val Gly Arg Glu Asp Ser Ser Asn Asn
 305 310 315
 Phe Arg Ile Asn Val Trp Leu Ala Val Leu Gln Met Ile Gln Asp
 320 325 330
 Arg Pro Trp Leu Gly Ile Gly Pro Gly Asn Thr Ala Phe Asn Leu
 335 340 345
 Val Tyr Pro Leu Tyr Gln Gln Ala Arg Phe Thr Ala Leu Ser Ala
 350 355 360
 Tyr Ser Val Pro Leu Glu Val Ala Val Glu Gly Gly Leu Leu Gly
 365 370 375
 Leu Thr Ala Phe Ala Trp Leu Leu Leu Val Thr Ala Val Thr Ala
 380 385 390
 Val Arg Gln Val Ser Arg Leu Arg Arg Asp Arg Asn Pro Gln Ala
 395 400 405
 Phe Trp Leu Met Ala Ser Leu Ala Gly Leu Ala Gly Met Leu Gly
 410 415 420
 His Gly Leu Phe Asp Thr Val Leu Tyr Arg Pro Glu Ala Ser Thr
 425 430 435
 Leu Trp Trp Leu Cys Ile Gly Ala Ile Ala Ser Phe Trp Gln Pro
 440 445 450
 Gln Pro Ser Lys Gln Leu Pro Pro Glu Ala Glu His Ser Asp Glu
 455 460 465
 Lys Met

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1425
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ATGGTGTCTC CCATCTCTAT CTGGCGATCG CTGATGTTTG GCGGTTTTTC 50
 CCCCCAGGAA TGGGGCCGGG GCAGTGTGCT CCATCGTTTG GTGGGCTGGG 100
 GACAGAGTTG GATACAGGCT AGTGTGCTCT GCGCCCACTT CGAGGCATTG 150
 GGTACGGCTC TAGTGGCAAT AATTTTTATT GCGGCTCCCT TCACCTCCAC 200
 CACCATGTTG GGCATTTTAA TGCTGCTCTG TGGAGCCTTT TGGGCTCTGC 250
 TGACCTTTGC TGATCAACCA GGGAAGGGTT TGACTCCCAT CCATGTTTTA 300
 GTTTTTGCCT ACTGGTGCAT TTCGCGATC GCCGTGGGAT TTTCTCCGGT 350
 AAAAATGGCG GCGGCGTCGG GGTAGCGGAA ATTAACAGCT AATTTATGTC 400
 TGTTTCTACT GCGCGCGAGG TTATTGCAAA ACAACAATG GTTGAACCGG 450
 TTAGTAACCG TTGTTTTACT GGTAGGGCTA TTGGTGGGGA GTTACGGTCT 500
 GCGACAACAG GTGGACGGGG TAGAACAGTT AGCCACTTGG AATGACCCCA 550
 CCTCTACCTT GGCCAGGCC ACTAGGGTAT ATAGCTTTTT AGGTAATCCC 600
 AATCTCTTGG CGGCTTACCT GGTGCCCATG ACGGGTTTGA GCTTGAGTGC 650
 CCTGGTGGTA TGGCGACGGT GGTGGCCCAA ACTGCTGGGA GCAACCATGG 700
 TGATTGTAA CCTACTCTGT CTCTTTTTAA CCCAGAGCCG GGGCGGTTGG 750
 CTAGCAGTGC TGGCCCTGGG AGCTACCTTC CTGGCCCTTT GTTACTTCTG 800
 GTGGTTACCC CAATTACCCA AATTTTGGCA ACGGTGGTCT TTGCCCTTGG 850
 CGATCGCCGT GCGGTTATA TTAGTGCGG GAGCGTTGAT TCGGTTGGAA 900
 CCGATTTCGAC TCAGGGCCAT GAGCATTTTT GCTGGGCGGG AAGACAGCAG 950
 TAATAATTTC CGCATCAATG TTTGGGAAGG GGTAAAAGCC ATGATCCGAG 1000
 CCCGCCCTAT CATTGGCATT GGCCAGGTA ACGAAGCCTT TAACCAAATT 1050
 TATCCTTACT ATATGCGGCC CCGCTTCACC GCCCTGAGTG CCTATTCCAT 1100
 TTACCTAGAA ATTTTGGTGG AAACGGGTGT AGTTGGTTTT ACCTGTATGC 1150
 TCTGGCTGTT GGCCGTTACC CTAGGCAAAG GCGTAGAACT GGTTAAACGC 1200
 TGTCGCCAAA CCCTCGCCCC GGAAGGCATC TGGATTATGG GGGCTTTAGC 1250
 GGCGATCATC GGTGTGTGG TCCACGGCAT GGTAGATACA GTCTGGTACC 1300

GTCCCCCGGT GAGCACTTTG TGGTGGTTGC TAGTGGCCAT TGTGCTAGT 1350
 CAGTGGGCCA GCGCCCAGGC CCGTTTGGAG GCCAGTAAAG AAGAAAATGA 1400
 GGACAAACCT CTTCTTGCTT CATAA 1425

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 474
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Met	Val	Ser	Pro	Ile	Ser	Ile	Trp	Arg	Ser	Leu	Met	Phe	Gly	Gly	5	10	15
Phe	Ser	Pro	Gln	Glu	Trp	Gly	Arg	Gly	Ser	Val	Leu	His	Arg	Leu	20	25	30
Val	Gly	Trp	Gly	Gln	Ser	Trp	Ile	Gln	Ala	Ser	Val	Leu	Trp	Pro	35	40	45
His	Phe	Glu	Ala	Leu	Gly	Thr	Ala	Leu	Val	Ala	Ile	Ile	Phe	Ile	50	55	60
Ala	Ala	Pro	Phe	Thr	Ser	Thr	Thr	Met	Leu	Gly	Ile	Phe	Met	Leu	65	70	75
Leu	Cys	Gly	Ala	Phe	Trp	Ala	Leu	Leu	Thr	Phe	Ala	Asp	Gln	Pro	80	85	90
Gly	Lys	Gly	Leu	Thr	Pro	Ile	His	Val	Leu	Val	Phe	Ala	Tyr	Trp	95	100	105
Cys	Ile	Ser	Ala	Ile	Ala	Val	Gly	Phe	Ser	Pro	Val	Lys	Met	Ala	110	115	120
Ala	Ala	Ser	Gly	Leu	Ala	Lys	Leu	Thr	Ala	Asn	Leu	Cys	Leu	Phe	125	130	135
Leu	Leu	Ala	Ala	Arg	Leu	Leu	Gln	Asn	Lys	Gln	Trp	Leu	Asn	Arg	140	145	150
Leu	Val	Thr	Val	Val	Leu	Leu	Val	Gly	Leu	Leu	Val	Gly	Ser	Tyr	155	160	165
Gly	Leu	Arg	Gln	Gln	Val	Asp	Gly	Val	Glu	Gln	Leu	Ala	Thr	Trp	170	175	180
Asn	Asp	Pro	Thr	Ser	Thr	Leu	Ala	Gln	Ala	Thr	Arg	Val	Tyr	Ser	185	190	195
Phe	Leu	Gly	Asn	Pro	Asn	Leu	Leu	Ala	Ala	Tyr	Leu	Val	Pro	Met	200	205	210
Thr	Gly	Leu	Ser	Leu	Ser	Ala	Leu	Val	Val	Trp	Arg	Arg	Trp	Trp	215	220	225
Pro	Lys	Leu	Leu	Gly	Ala	Thr	Met	Val	Ile	Val	Asn	Leu	Leu	Cys	230	235	240
Leu	Phe	Phe	Thr	Gln	Ser	Arg	Gly	Gly	Trp	Leu	Ala	Val	Leu	Ala	245	250	255
Leu	Gly	Ala	Thr	Phe	Leu	Ala	Leu	Cys	Tyr	Phe	Trp	Trp	Leu	Pro	260	265	270
Gln	Leu	Pro	Lys	Phe	Trp	Gln	Arg	Trp	Ser	Leu	Pro	Leu	Ala	Ile	275	280	285
Ala	Val	Ala	Val	Ile	Leu	Gly	Gly	Gly	Ala	Leu	Ile	Ala	Val	Glu	290	295	300
Pro	Ile	Arg	Leu	Arg	Ala	Met	Ser	Ile	Phe	Ala	Gly	Arg	Glu	Asp	305	310	315
Ser	Ser	Asn	Asn	Phe	Arg	Ile	Asn	Val	Trp	Glu	Gly	Val	Lys	Ala	320	325	330
Met	Ile	Arg	Ala	Arg	Pro	Ile	Ile	Gly	Ile	Gly	Pro	Gly	Asn	Glu	335	340	345

Ala Phe Asn Gln Ile Tyr Pro Tyr Tyr Met Arg Pro Arg Phe Thr
 350 355 360
 Ala Leu Ser Ala Tyr Ser Ile Tyr Leu Glu Ile Leu Val Glu Thr
 365 370 375
 Gly Val Val Gly Phe Thr Cys Met Leu Trp Leu Leu Ala Val Thr
 380 385 390
 Leu Gly Lys Gly Val Glu Leu Val Lys Arg Cys Arg Gln Thr Leu
 395 400 405
 Ala Pro Glu Gly Ile Trp Ile Met Gly Ala Leu Ala Ala Ile Ile
 410 415 420
 Gly Leu Leu Val His Gly Met Val Asp Thr Val Trp Tyr Arg Pro
 425 430 435
 Pro Val Ser Thr Leu Trp Trp Leu Leu Val Ala Ile Val Ala Ser
 440 445 450
 Gln Trp Ala Ser Ala Gln Ala Arg Leu Glu Ala Ser Lys Glu Glu
 455 460 465
 Asn Glu Asp Lys Pro Leu Leu Ala Ser
 470

(2) INFORMATION FOR SEQ ID NO:6:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 31
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:
 GGGCTAGCCG CGATCGCGGC CTATTGGGCC C 31

(2) INFORMATION FOR SEQ ID NO:7:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:
 GGGCTAGGGA TCGCGCCTAT TGGGCC 27

(2) INFORMATION FOR SEQ ID NO:8:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 26
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:
 GGGCTCAGAT CGCGCCTATT GGGCCC 26

(2) INFORMATION FOR SEQ ID NO:9:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 11
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:
 Gly Leu Ala Ala Ile Ala Ala Tyr Trp Ala Leu
 5 10

TCGCGCCTAT TGGGCC 27